Approximating the Permanent via Nonabelian Determinants

Cristopher Moore* Alexander Russell†

June 9, 2009

Abstract

Since the celebrated work of Jerrum, Sinclair, and Vigoda, we have known that the permanent of a $\{0,1\}$ matrix can be approximated in randomized polynomial time by using a rapidly mixing Markov chain to sample perfect matchings of a bipartite graph. A separate strand of the literature has pursued the possibility of an alternate, *algebraic* polynomial-time approximation scheme. These schemes work by replacing each 1 with a random element of an algebra $\mathcal{A}$, and considering the determinant of the resulting matrix.

In the case where $\mathcal{A}$ is noncommutative, this determinant can be defined in several ways. We show that for estimators based on the conventional determinant, the critical ratio of the second moment to the square of the first—and therefore the number of trials we need to obtain a good estimate of the permanent—is $(1 + O(1/d))^n$ when $\mathcal{A}$ is the algebra of $d \times d$ matrices. These results can be extended to group algebras, and semi-simple algebras in general.

We also study the *symmetrized* determinant of Barvinok, showing that the resulting estimator has small variance when d is large enough. However, if d is constant—the only case in which an efficient algorithm is known—we show that the critical ratio exceeds $2^n/n^{O(d)}$. Thus our results do not provide a new polynomial-time approximation scheme for the permanent. Indeed, they suggest that the algebraic approach to approximating the permanent faces significant obstacles.

We obtain these results using diagrammatic techniques in which we express matrix products as contractions of tensor products. When these matrices are random, in either the Haar measure or the Gaussian measure, we can evaluate the trace of these products in terms of the cycle structure of a suitably random permutation. In the symmetrized case, our estimates are then derived by a connection with the character theory of the symmetric group.

1 Introduction

The *permanent* of an $n \times n$ matrix A is $\operatorname{perm} A = \sum_{\pi \in S_n} \prod_{i=1}^n A_{i,\pi i}$, where S_n denotes the group of permutations of n objects. If $A_{ij} \in \{0,1\}$ for all i, j, we can also write $\operatorname{perm} A = |\{\pi \in S_n \mid \pi \vdash A\}|$ where $\pi \vdash A$ denotes the following relation,

$$\pi \vdash A \Leftrightarrow A_{i,\pi i} = 1 \text{ for all } i\,.$$

Computing the permanent of a $\{0,1\}$ matrix is $\#P$-complete. Therefore, we cannot expect to compute it efficiently without startling complexity-theoretic consequences, including the collapse of the polynomial hierarchy [Val79, Tod91].

*moore@santafe.edu, Department of Computer Science, University of New Mexico and Santa Fe Institute
†acr@cse.uconn.edu, Department of Computer Science and Engineering, University of Connecticut

On the other hand, Godsil and Gutman [GG81] pointed out the following charming fact. If we define the matrix-valued random variable M so that $M_{ij} = \rho_{ij} A_{ij}$, where the ρ_{ij} are chosen independently and uniformly from $\{\pm 1\}$, and define $X = (\det M)^2$, then it is easy to check that X is an estimator for $\operatorname{perm} A$, which is to say that $\mathbb{E}[X] = \operatorname{perm} A$. Since $\det M$ can be computed efficiently, so can X. This suggests a natural randomized approximation algorithm for the permanent: average a family of independent samples of X.

The quality of this approximation can be controlled by determining the variance of X. If X_t denotes the average of X over t independent trials, then Chebyshev's inequality shows that, in order for X_t to yield an approximation of $\mathbb{E}[X]$ within a factor $\alpha = O(1)$ with probability $\Omega(1)$, the number of trials we need is at most

$$t \sim \frac{\mathsf{Var}\, X}{\mathbb{E}[X]^2} \leq \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2}\,.$$

Following [CRS03], we refer to this quantity as the *critical ratio* of the estimator. Karmarkar, Karp, Lipton, Lovász, and Luby [KKL+93] showed, unfortunately, that the critical ratio is $3^{n/2}$ in the worst case, ignoring $\operatorname{poly}(n)$ factors. Then again, they showed that we can decrease this to $2^{n/2}$ by drawing ρ_{ij} uniformly from the unit circle in the complex plane, or simply from the cube roots of unity, instead of $\{\pm 1\}$. Barvinok [Bar99] obtained a more concentrated estimator by drawing ρ_{ij} from normal distributions over $\mathbb{R}$, $\mathbb{C}$, and the quaternions $\mathbb{H}$.

This raises the interesting possibility that, by choosing the ρ_{ij} from the right set of algebraic objects, we might be able to reduce the critical ratio to $\mathrm{e}^{o(n)}$, or even to $\operatorname{poly}(n)$, resulting in a subexponential or polynomial-time algorithm. One exciting result in this direction is due to Chien, Rasmussen, and Sinclair [CRS03], who showed that certain determinants defined over the Clifford algebra CL_k with k generators give estimators where the critical ratio is $(1 + O(2^{-k/2}))^{n/2}$. In the case of the quaternion group, where $k = 3$, they gave a polynomial-time algorithm for a type of determinant where the critical ratio grows as $(3/2)^{n/2}$. This is currently the best known critical ratio for an algebraic estimator which can be computed efficiently. Sadly, however, for larger k we do not know how to compute these determinants in polynomial time.

These results can be given a uniform presentation by defining a notion of determinant for a matrix M over an associative algebra $\mathcal{A}$. The traditional Cayley determinant is then

$$\det M = \sum_{\alpha \in S_n} (-1)^\alpha \prod_{i=1}^{n} M_{i,\alpha i}\,, \tag{1}$$

where $(-1)^\alpha$ denotes the sign of the permutation α. Note that $\det M$ takes values in $\mathcal{A}$. If $\mathcal{A}$ is noncommutative, however, the determinant as defined in (1) may depend on the order in which the product is taken. As written, each traversal $M_{i,\alpha i}$ is ordered from the top row to the bottom row; we could just as easily order them from the left column to the right. This introduces some arbitrariness to the definition, and appears to complicate the problem of computing such determinants, even when the algebra $\mathcal{A}$ has small dimension [Nis91].

One natural remedy is to remove this order dependence by forcibly symmetrizing each product appearing in (1). This gives the following *symmetrized determinant*,

$$\operatorname{sdet} M = \frac{1}{n!} \sum_{\alpha, \alpha' \in S_n} (-1)^{\alpha' \alpha^{-1}} \prod_{i=1}^{n} M_{\alpha i, \alpha' i}\,. \tag{2}$$

Observe that sdet is obtained by symmetrizing each product appearing in (1). This definition is due to Barvinok [Bar00], who showed that if $\mathcal{A}$ has dimension m, the symmetrized determinant can be computed in time $O(n^{m+O(1)})$. In contrast, no efficient algorithm is currently known for the unsymmetrized Cayley determinant (1), even when the dimension of $\mathcal{A}$ is constant.

We focus on the algebra $\mathcal{A}_d$, consisting of all $d \times d$ matrices over $\mathbb{C}$. We remark that any finite dimensional C^*-algebra,[1] which appear to be the natural settings for such approximations, are *semi-simple*, meaning that they can be decomposed as a direct product of algebras of the form $\mathcal{A}_d$. In particular, all group algebras and the Clifford algebras studied in [CRS03] have this property. It follows that many of our results, especially lower bounds on the critical ratio, carry over easily to estimators based on suitable distributions in semisimple algebras.

Now, given a matrix A with entries in $\{0,1\}$, define $M_{ij} = \rho_{ij} A_{ij}$, where the ρ_{ij} are independently random $d \times d$ matrices. (We focus on $\{0,1\}$ matrices, but we can let the A_{ij} be arbitrary nonnegative reals by taking $M_{ij} = \rho_{ij}\sqrt{A_{ij}}$.) Since the M_{ij} take values in $\mathcal{A}_d$, then so do $\det M$ and $\operatorname{sdet} M$. There are several ways to turn these matrix-valued determinants into real-valued estimators for the permanent of a real-valued matrix A. As mentioned above, most of the existing literature has focused on the Frobenius norm of these determinants. For technical reasons, we focus first on the absolute value squared of their trace. This gives us two estimators,

$$X = |\mathbf{tr} \det M|^2 \quad \text{and} \quad X_s = |\mathbf{tr} \operatorname{sdet} M|^2 \ .$$

Note that these are random $\mathbb{R}$-valued variables depending on the ρ_{ij}. We will then address the Frobenius estimators,

$$X_{\text{Frob}} = \|\det M\|^2 \quad \text{and} \quad X_{\text{Frob},s} = \|\operatorname{sdet} M\|^2 \ .$$

As an additional degree of freedom, we can draw ρ_{ij} according to two different distributions on $\mathcal{A}_d$. The *Haar measure* is the uniform distribution over unitary matrices. In the *Gaussian measure*, each entry of ρ_{ij} is drawn independently from the Gaussian distribution on $\mathbb{C}$ with mean 0 and variance $1/d$: that is, its real and imaginary parts are drawn independently from the Gaussian distribution on $\mathbb{R}$ with mean 0 and variance $1/(2d)$, $p(x) = \mathrm{e}^{-x^2}/\sqrt{\pi}$.

Our main contribution is given by the following theorems.

Theorem 1. *For both the Haar and Gaussian measures, in the unsymmetrized case we have*

$$\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} = \left(1 + O\left(\frac{1}{d}\right)\right)^n \ . \tag{3}$$

In the symmetrized case,

$$\frac{\mathbb{E}[X_s^2]}{\mathbb{E}[X_s]^2} \leq 2^{2n}\, n^{-d+O(1)} \ \textit{if } d = O(1) \tag{4}$$

and more generally,

$$\frac{\mathbb{E}[X_s^2]}{\mathbb{E}[X_s]^2} = O\left(\mathrm{e}^{4n^2/d}\right) \ . \tag{5}$$

Additionally, we establish lower bounds on the critical ratio $\mathbb{E}[X_s^2]/\,\mathbb{E}[X_s]^2$.

[1] A C^*-algebra is an algebra over $\mathbb{R}$ or $\mathbb{C}$ possessing a norm $\|\cdot\|$ and an involution operator $\cdot^*$ consistent in the sense that $\|x^*x\|^2 = \|x\|^2$. See, e.g., [Con00, §1] for a complete definition.

Theorem 2. *Let A be the $n \times n$ identity matrix and d a constant. Then*

$$\frac{\mathbb{E}[X_s^2]}{\mathbb{E}[X_s]^2} = \Omega\left(\frac{2^n}{n^d}\right) \quad \textit{and} \quad \frac{\mathbb{E}[X_s^2]}{\mathbb{E}[X_s]^2} = \left(1 - O\left(\frac{1}{d}\right)\right)^n \Omega\left(\frac{2^n}{n^d}\right),$$

when the ρ_{ij} are distributed according to the Gaussian or Haar measure respectively.

Finally, we show the critical ratio differs by at most d^4 for the Frobenius estimators than for those given by the square of the trace:

Theorem 3.

$$\frac{1}{d^4}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} \leq \frac{\mathbb{E}[X_{\text{Frob}}^2]}{\mathbb{E}[X_{\text{Frob}}]^2} \leq d^4 \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2}, \tag{6}$$

and similarly for $X_{\text{Frob},s}$.

These results give a somewhat frustrating outlook. The critical ratio for the unsymmetrized estimator behaves very well, becoming more and more mildly exponential as d increases, much like the Clifford group estimator of [CRS03]. However, we do not know how to compute these estimators efficiently. On the other hand, we can compute the symmetrized estimator if d is constant [Bar00], but our results show that its critical ratio does not decrease appreciably until d is roughly n^2.

Barvinok [Bar00] suggested that the estimators $X_{\text{Frob},s}$ might become asymptotically concentrated when d is large, but constant. Specifically, he made the following conjecture (where we have weakened the lower bound, specialized to $\{0,1\}$ matrices, and changed the notation to fit our purposes):

Conjecture 1. *If A is an $n \times n$ matrix with entries in $\{0,1\}$, let $M(A)$ be the matrix $M_{ij} = \rho_{ij} A_{ij}$, where each ρ_{ij} is chosen independently from the Gaussian distribution on $\mathcal{A}_d$. Define $M(\mathbb{1})$ similarly, where $M_{ij} = \rho_{ij}\delta_{ij}$. Then there is a sequence of constants γ_d, where $\lim_{d\to\infty}\gamma_d = 1$, such that for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \Pr\left[(\gamma_d + \epsilon)^{-n} \operatorname{perm} A \leq \frac{\|M(A)\|^2}{\|M(\mathbb{1})\|^2} \leq (\gamma_d + \epsilon)^n \operatorname{perm} A\right] = 1.$$

Our results do not address Conjecture 1 directly, However, given Chebyshev's inequality, it is very natural to consider the following stronger conjecture, which would imply Conjecture 1:

Conjecture 2. *There is a sequence of constants θ_d, where $\lim_{d\to\infty}\theta = 1$, such that for any $n \times n$ matrix A, the critical ratio of the estimator $X_{\text{Frob},s} = \|M(A)\|^2$ obeys*

$$\frac{\mathbb{E}[X_{\text{Frob},s}^2]}{\mathbb{E}[X_{\text{Frob},s}]^2} \leq \theta_d^n.$$

Sadly, Theorems 2 and 3 imply that Conjecture 2 is false. It is still conceivable that Conjecture 1 is true, but it seems that any proof of it would have to bound higher moments of the estimator: the first and second moments alone do not scale in a way that gives concentration.

The remainder of the paper is organized as follows. In Section 2, we calculate the expectations of these estimators, showing that they are each a constant a_d times the permanent, and computing the constant explicitly using a diagrammatic technique. In Sections 3 and 4, we bound their second moments using the same technique, proving Theorems 1 and 2. In Section 5, we relate the critical ratio for the Frobenius estimators to the trace-squared estimators, proving Theorem 3. Finally, in Section 6 we discuss the implications of this theorem, and the remaining barriers to an algebraic approximation scheme for the permanent.

2 The expectation

Before we proceed, we write the following expansions for these estimators, which we will find useful for calculating their expectations and second moments:

$$X = \sum_{\alpha,\beta \vdash A} (-1)^{\alpha\beta} \left(\mathbf{tr} \prod_i \rho_{i,\alpha i} \right) \left(\mathbf{tr} \prod_i \rho^*_{i,\beta i} \right) \tag{7}$$

$$X_s = \sum_{\kappa,\lambda \vdash A} (-1)^{\kappa\lambda} \, \mathbb{E}_{\alpha,\beta} \left(\mathbf{tr} \prod_i \rho_{\alpha i,\kappa\alpha i} \right) \left(\mathbf{tr} \prod_i \rho^*_{\beta i,\lambda\beta i} \right) . \tag{8}$$

Since $\mathbb{E}[\rho_{ij}] = 0$, any term in which some ρ_{ij} appears only once will have zero expectation. Then the cross-terms in the expansion (7) are zero in expectation except when $\alpha = \beta$, so

$$\mathbb{E}[X] = \sum_{\alpha \vdash A} \mathbb{E} \left(\mathbf{tr} \prod_i \rho_{i,\alpha i} \right) \left(\mathbf{tr} \prod_i \rho^*_{i,\alpha i} \right) = \sum_{\alpha \vdash A} \mathbb{E} \left| \mathbf{tr} \prod_i \rho_{i,\alpha i} \right|^2 = \operatorname{perm} A \, . \tag{9}$$

Here we used the use the following fact, which is an easy exercise: if σ is the product of n independent random matrices, chosen from the Haar measure or the Gaussian measure, then $\mathbb{E} \left| \mathbf{tr} \, \sigma \right|^2 = 1$.

Similarly, the only terms in (8) that contribute to $\mathbb{E}[X_s]$ are those where $\lambda = \kappa$, so that each ρ_{ij} appears twice or not at all. Thus

$$\mathbb{E}_{\{\rho_{ij}\}}[X_s] = \sum_{\kappa \vdash A} \mathbb{E}_{\{\sigma_i\}} \, \mathbb{E}_{\alpha,\beta} \left(\mathbf{tr} \prod_i \sigma_{\alpha i} \right) \left(\mathbf{tr} \prod_i \sigma^*_{\beta i} \right) = a_d \cdot \operatorname{perm} A$$

$$\text{where} \quad a_d = \mathbb{E}_{\{\sigma_i\}} \, \mathbb{E}_{\alpha,\beta} \left(\mathbf{tr} \prod_i \sigma_{\alpha i} \right) \left(\mathbf{tr} \prod_i \sigma^*_{\beta i} \right) . \tag{10}$$

A similar result for the Frobenius estimator $X_{\mathrm{Frob},s} = \|\operatorname{sdet} M\|^2$ appears as Theorem 4.3 in Barvinok [Bar00].

We can think of a_d as the expectation, over all pairs of permutations α, β, of the covariance between the trace of two products of the same n random matrices, where the products are taken in the orders given by α and β. This expectation clearly stays the same if we assume that α is the identity 1 and β is uniformly random, so we can write

$$a_d = \mathbb{E}_\beta \, \mathbb{E}_{\{\sigma_i\}} \left(\mathbf{tr} \prod_i \sigma_i \right) \left(\mathbf{tr} \prod_i \sigma^*_{\beta i} \right) .$$

We will evaluate these covariances using a diagrammatic approach. First, suppose we have n linear operators $\sigma_1, \ldots, \sigma_n$. The trace of their product is

$$(\sigma_1)^{i_1}_{i_2} (\sigma_2)^{i_2}_{i_3} \cdots (\sigma_n)^{i_n}_{i_1} .$$

Here we save ink by using the Einstein summation convention, in which any index that appears twice is automatically summed over. We can think of this product as a particular kind of internal trace of the tensor product

$$(\sigma_1 \otimes \cdots \otimes \sigma_n)^{i_1,\ldots,i_n}_{j_1,\ldots,j_n} = (\sigma_1)^{i_1}_{j_1} \cdots (\sigma_n)^{i_n}_{j_n} ,$$

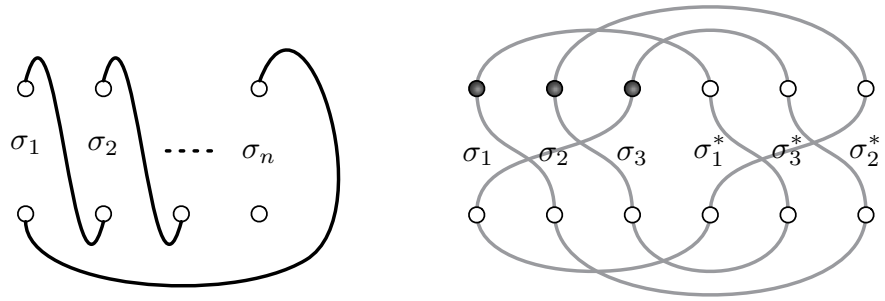


Figure 1: The trace of the matrix product $\sigma_1\sigma_2\cdots\sigma_n$ is a contraction of the tensor product $\sigma_1 \otimes \cdots \otimes \sigma_n$. Combining this with (13) shows that the covariance between the traces of two permuted products is given by d^{c-n} where c is the number of loops in a diagram like that on the right. In this case, the covariance between $\mathbf{tr}\,\sigma_1\sigma_2\sigma_3$ and $\mathbf{tr}\,\sigma_2\sigma_1\sigma_3$ is $1/d^2$, since $n = 3$ and $c = 1$.

where we contract the index i_t with $j_{(i+1) \bmod n}$ for each i. We draw this on the left-hand side of Fig. 1. Then if $n = 3$, say, and β is the the transposition $(2\,3)$, the covariance

$$\mathbb{E}_{\sigma_1,\sigma_2,\sigma_3} (\mathbf{tr}\,\sigma_1\sigma_2\sigma_3) (\mathbf{tr}\,\sigma_1\sigma_3\sigma_2)^*$$

becomes a certain contraction of the tensor product of three independent and identical expectations,

$$\mathbb{E}_{\sigma_1} [\sigma_1 \otimes \sigma_1^*] \otimes \mathbb{E}_{\sigma_2} [\sigma_2 \otimes \sigma_2^*] \otimes \mathbb{E}_{\sigma_3} [\sigma_3 \otimes \sigma_3^*] \,. \tag{11}$$

The following lemma is well-known; we prove it in Appendix B for completeness.

Lemma 4. *If σ is chosen according to the Haar measure or the Gaussian measure, then*

$$\mathbb{E}_\sigma [\sigma \otimes \sigma^*]^{ik}_{j\ell} = \frac{1}{d}\,\delta^{ik}\delta_{j\ell}\,. \tag{12}$$

We can represent (12) diagrammatically as a "cupcap,"

$$\mathbb{E}_\sigma [\sigma \otimes \sigma^*] = \frac{1}{d} \bigcup\!\!\bigcap. \tag{13}$$

A tensor product such as (11) becomes three cupcaps side by side, and contracting it consists of connecting pairs of inputs and outputs until the diagram becomes closed. For instance, the expectation of $(\mathbf{tr}\,\sigma_1\sigma_2\sigma_3)\,(\mathbf{tr}\,\sigma_1\sigma_3\sigma_2)^*$ corresponds to the diagram on the right-hand side of Fig. 1. Here we have drawn the cupcaps on the top and bottom of the diagram (between corresponding indices of σ_i and σ_i*) and the connections between them in the interior.

When we evaluate the trace of this diagram, each of the n cupcaps introduces a factor of $1/d$ according to (13), and each loop in the diagram corresponds to an index which can be set independently to any value between 1 and d. So, the diagram evaluates to d^{c-n} where c is the number of loops. In this case $n = 3$ and $c = 1$, and the covariance is $1/d^2$.

More generally, we can write the covariance between $\mathbf{tr}\prod_i \sigma_i$ and $\mathbf{tr}\prod_i \sigma_{\beta i}$ as a function of β as follows. The cupcaps match the upper indices of the σs in the first product to those of the second

product according to β, and the lower indices of the second product to those of the first product according to β^{-1}. If r denotes the rotation $(1\,2\,\cdots\,n)$, which "weaves" the σs together and takes the trace of their product, then following the diagram around gives a permutation on (say) the upper n indices of the first product (darkened in Fig. 1) equal to the commutator $[\beta, r] = \beta r \beta^{-1} r^{-1}$. Each loop in the diagram corresponds to a cycle in this permutation. So, we have

$$a_d = \frac{1}{d^n}\, \mathbb{E}_\beta\, d^{c([\beta,r])} \tag{14}$$

where $c(\pi)$ denotes the number of cycles in a permutation π. Note that we always have

$$a_d \geq \frac{n}{n!}\,. \tag{15}$$

This follows because, with probability $n/n!$, a uniformly random β is one of the n powers of r. In that case $[\beta, r] = 1$, and $d^{c([\beta,r])} = d^n$. It can be shown that this bound is tight when $d = \omega(n^2)$.

The expectation (14) can be viewed as the inner product of P_n, the uniform distribution over the conjugacy class $[r] = \{\pi^{-1} r \pi \mid \pi \in S_n\}$, and the function $d^{c(\cdot)}$, both of which are *class functions*—invariant under conjugation. Below, we show that these can be expanded in terms of the characters of the group S_n and analyzed using the Littlewood-Richardson rule; this yields an exact expression for a_d:

Lemma 5.

$$\textit{If } d \leq n,\ a_d = \frac{1}{d^n}\binom{n+d}{n+1}. \quad \textit{If } d > n,\ a_d = \frac{1}{d^n}\left(\binom{n+d}{n+1} - \binom{d}{n+1}\right). \tag{16}$$

Proof. First, note that the function $d^{c(\pi)}$ is a *class function*, i.e., one which is invariant under conjugation. Therefore, in $\mathbb{E}_\beta\, d^{c([\beta,r])}$ we can replace $[\beta, r]$ with $\zeta[\beta, r]\zeta^{-1}$ where β and ζ are uniformly random. Since

$$\zeta[\beta, r]\zeta^{-1} = \zeta\beta r \beta^{-1} r^{-1} \zeta^{-1} = \big((\zeta\beta) r (\zeta\beta)^{-1}\big)\big(\zeta r^{-1}\zeta^{-1}\big)\,,$$

we can treat this as the expectation of $d^{c(\pi)}$ where π is the product of two uniformly random elements of $[r]$, the conjugacy class consisting of cycles of length n. In other words, if $P_n : S_n \to \mathbb{R}$ is the uniform distribution on the conjugacy class of n-cycles, then

$$a_d = \frac{1}{d^n} \sum_\pi (P_n * P_n)(\pi)\, d^{c(\pi)} \tag{17}$$

where $P_n * P_n$ is the convolution of P_n with itself,

$$(P_n * P_n)(\pi) = \sum_{\eta \in S_n} P_n(\eta)\, P_n(\eta^{-1}\pi)\,.$$

We will view (17) as an inner product over S_n,

$$a_d = \frac{n!}{d^n} \left\langle P_n * P_n, d^{c(\cdot)} \right\rangle\,, \tag{18}$$

where the inner product over a group G of two functions $f_1, f_2 : G \to \mathbb{C}$ is defined as

$$\langle f_1, f_2 \rangle = \frac{1}{|G|} \sum_{g \in G} f_1(g)^* f_2(g) \,.$$

To evaluate (18), we will expand P and $d^{c(\cdot)}$ in the Fourier basis, as a sum of irreducible characters of S_n. Recall that the characters of a finite group are orthonormal under the inner product above and, additionally, convolution is transformed to pointwise product in the Fourier basis. In short, for two characters χ and ψ,

$$\chi * \psi = \begin{cases} \frac{|G|}{\chi(1)}\,\chi & \text{if } \chi = \psi, \\ 0 & \text{if } \chi \neq \psi, \end{cases} \quad \text{and} \quad \langle \chi, \psi \rangle = \begin{cases} 1 & \text{if } \chi = \psi, \\ 0 & \text{if } \chi \neq \psi. \end{cases} \tag{19}$$

Each character of the symmetric group is associated with a Young diagram, i.e., a partition $\lambda_1 \geq \lambda_2 \geq \cdots$ where $\sum_i \lambda_i = n$. In light of the Murnaghan-Nakayama rule (Lemma 10 of Appendix A), the uniform distribution P_n over the conjugacy class $[r]$ is supported solely on *hooks*, i.e., Young diagrams consisting of a single ribbon of size n. Let Λ_t denote the hook of height $t+1$, in which $\lambda_1 = n - t$ for some $0 \leq t < n$ and $\lambda_i = 1$ for $1 < i \leq t+1$. Let χ_t denote the corresponding character. Then $\dim \chi_t = \chi_t(1) = \binom{n-1}{t}$ and, again appealing to Lemma 10, we have $\chi_t([r]) = (-1)^t$. Applying (19) then gives

$$\langle P_n, \chi_t \rangle = \frac{(-1)^t}{n!} \quad \text{and} \quad \langle P_n * P_n, \chi_t \rangle = \frac{1}{n!} \frac{1}{\binom{n-1}{t}} \,. \tag{20}$$

To calculate the inner product $\left\langle d^{c(\cdot)}, \chi_t \right\rangle$, consider the following combinatorial representation of S_n. Let Σ be the set of strings of length n over the alphabet $\{1, \ldots, d\}$, and let S_n act on Σ in the natural way, by permuting the symbols in a given string. Given a permutation π, the character $\chi_\Sigma(\pi)$ is the number of strings fixed by π. Since each of π's cycles can be given an independent label in $\{1, \ldots, d\}$, we have $\chi_\Sigma(\pi) = d^{c(\pi)}$.

It follows that $\left\langle d^{c(\cdot)}, \chi_t \right\rangle = \langle \chi_t, \chi_\Sigma \rangle$, the number of copies of Λ_t appearing in the decomposition of Σ into irreducible representations. To find this, we first decompose Σ into a direct sum of combinatorial representations $\Sigma_{(n_1,\ldots,n_d)}$, consisting of strings where i appears n_i times for each $i \in \{1, \ldots, d\}$. Then $\left\langle \chi_{(n_1,\ldots,n_d)}, \chi_t \right\rangle$ is given by a *Kostka number*, defined as follows. First, sort the n_i in decreasing order so that they form a Young diagram N. Then $K_N^{\Lambda_t} = \left\langle \chi_{(n_1,\ldots,n_d)}, \chi_t \right\rangle$ is the number of semistandard tableaux of shape Λ_t and content N: that is, the number of ways to fill Λ_t with n_i is for each $i \in \{1, \ldots, d\}$, where each row is nondecreasing and where each column is strictly increasing.

Since Λ_t is a hook, to specify a semistandard tableau with a given content it suffices to specify the content of the leftmost column. Since this column must be strictly increasing, its $t+1$ entries must be distinct. If N has k rows, i.e., if $n_i \neq 0$ for k values of i, then the first one must appear in the top cell, but the remaining t cells can be chosen arbitrarily. Thus $K_N^{\Lambda_t} = \binom{k-1}{t}$, and is 0 if $t \geq k$. There are $\binom{d}{k}\binom{n-1}{k-1}$ partitions $(n_1, \ldots, n_d)$ with k nonzero n_i. Since $1 \leq k \leq \min(d, n)$, summing over k then gives

$$\left\langle d^{c(\pi)}, \chi_t \right\rangle = \sum_k \binom{d}{k} \binom{n-1}{k-1} K_N^{\Lambda_t} = \sum_{k=1}^{\min(d,n)} \binom{d}{k} \binom{n-1}{k-1} \binom{k-1}{t} \,. \tag{21}$$

We can now calculate the inner product $\left\langle P_n, d^{c(\cdot)} \right\rangle$. Combining (20) and (21) and summing over t, we have

$$n! \left\langle P_n, d^{c(\cdot)} \right\rangle = n! \sum_{t=0}^{n-1} \langle P * P, \chi_t \rangle \left\langle d^{c(\cdot)}, \chi_t \right\rangle = \sum_{t=0}^{n-1} \sum_{k=1}^{\min(d,n)} \binom{d}{k} \binom{n-1}{k-1} \binom{k-1}{t} \Big/ \binom{n-1}{t}$$
$$= \sum_{k=1}^{\min(d,n)} \binom{d}{k} \sum_{t=0}^{k-1} \binom{n-t-1}{n-k} = \sum_{k=1}^{\min(d,n)} \binom{d}{k} \binom{n}{k-1} = \binom{n+d}{n+1} - \binom{d}{n+1},$$

where $\binom{d}{n+1} = 0$ if $d \leq n$. Combining this with (18) completes the proof. □

3 The second moment in the unsymmetrized case

Squaring (7)—and, for aesthetic reasons, placing the conjugated ρs in the second half of the expression and changing the names of the permutations—gives

$$X^2 = \sum_{\kappa,\lambda,\mu,\nu \vdash A} (-1)^{\kappa\lambda\mu\nu} \left(\mathbf{tr} \prod_i \rho_{i,\kappa i} \right) \left(\mathbf{tr} \prod_i \rho_{i,\lambda i} \right) \left(\mathbf{tr} \prod_i \rho^*_{i,\mu i} \right) \left(\mathbf{tr} \prod_i \rho^*_{i,\nu i} \right). \tag{22}$$

Now we take the expectation over the ρ_{ij}. As before, the only terms of this sum that contribute to this expectation are those in which each ρ_{ij} appears an even number of times. Moreover, each ρ_{ij} must appear an equal number of times conjugated (in the first and second products) and unconjugated (in the third and fourth products), since $\mathbb{E}_\sigma[\sigma \otimes \sigma] = 0$. In the Gaussian measure, this is because $\mathbb{E}[(\sigma^i_j)^2] = 0$ if σ^i_j is chosen from the Gaussian distribution on $\mathbb{C}$. In the Haar measure, the same thing is true because the tensor square $\sigma \otimes \sigma$ of the defining representation of $\mathsf{U}(d)$ contains no copies of the trivial representation.

For each term of (22), associated with a tuple $(\kappa, \lambda, \mu, \nu)$, we express the total number of occurrences of each ρ_{ij} with an $n \times n$ matrix C_{ij}. In light of the discussion above, for the terms that contribute to the second moment we have $C_{ij} = 0$ if $A_{ij} = 0$, $C_{ij} \in \{2, 4\}$ if $A_{ij} = 1$, and $\sum_i C_{ij} = \sum_j C_{ij} = 2n$. We will denote these conditions as $C \vdash A$. As in [KKL+93, CRS03], we think of C as a "double cycle cover" of the bipartite graph described by A. This graph has n vertices on either side, and an edge between the ith vertex on the left and the jth vertex on the right if and only if $A_{ij} = 1$. Each vertex has degree 2 or 4 in C. Thus C consists of cycles where each edge is covered twice, and possibly some isolated edges which are covered four times.

We then write the second moment as a sum, over all C, of the quadruples such that $(\kappa, \lambda, \mu, \nu) \vdash C$, where this denotes the following relation:

$$\begin{aligned}(\kappa, \lambda, \mu, \nu) \vdash C \Leftrightarrow\ & \pi \vdash A \text{ for all } \pi \in \{\kappa, \lambda, \mu, \nu\} \\ & \text{and } |\{\pi \in \{\kappa, \lambda\} \mid j = \pi i\}| = |\{\pi \in \{\mu, \nu\} \mid j = \pi i\}| \text{ for all } i, j \in \{1, \ldots, n\}, \\ & \text{and } |\{\pi \in \{\kappa, \lambda, \mu, \nu\} \mid j = \pi i\}| = C_{ij} \text{ for all } i, j \in \{1, \ldots, n\}.\end{aligned}$$

In our discussion below, we will treat each $(\kappa, \lambda, \mu, \nu)$ as a "coloring" of C. Each double edge is colored (κ, μ), (κ, ν), (λ, μ), or (λ, ν), indicating some pair ρ_{ij}, ρ^*_{ij} appearing in the first and third products, or the first and fourth, and so on. Each cycle in C must alternate between (κ, μ) and (λ, ν) or between (κ, ν) and (λ, μ). The isolated edges in C bear all four colors, indicating that some ρ_{ij} appears in all four products. We observe that for those tuples that contribute to the second moment, the parity $(-1)^{\kappa\lambda\mu\nu}$ is always 1.

Lemma 6. *If $(\kappa, \lambda, \mu, \nu) \vdash C$ for some C, then $(-1)^{\kappa\lambda\mu\nu} = 1$.*

Proof. Observe that $(-1)^{\kappa\lambda\mu\nu} = (-1)^{\pi}$ where $\pi = \kappa^{-1}\mu\lambda^{-1}\nu$. We claim that the constraints we describe above imply that $\pi = 1$. Consider a cycle c of C on the bipartite graph defined by A. We can view $\kappa, \lambda, \mu, \nu$ as one-to-one mappings from the n vertices on the left side to the n vertices on the right. If c alternates between (κ, μ) and (λ, ν), then restricting to the vertices on the left side of c we have $\kappa = \mu$ and $\lambda = \nu$. Similarly, if c alternates between (κ, ν) and (λ, μ), then restricting to these vertices gives $\kappa = \nu$ and $\lambda = \mu$. Finally, for an isolated edge we have $\kappa = \lambda = \mu = \nu$ when restricted to its left endpoint. In all cases we have $\kappa^{-1}\mu\lambda^{-1}\nu = 1$. □

Thus the second moment of the unsymmetrized estimator can be written

$$\mathbb{E}[X^2] = \sum_{C \vdash A} \sum_{(\kappa,\lambda,\mu,\nu) \vdash C} \mathbb{E}_{\{\rho_{ij}\}} \left(\mathbf{tr} \prod_i \rho_{i,\kappa i} \right) \left(\mathbf{tr} \prod_i \rho_{i,\lambda i} \right) \left(\mathbf{tr} \prod_i \rho^*_{i,\mu i} \right) \left(\mathbf{tr} \prod_i \rho^*_{i,\nu i} \right) . \tag{23}$$

Many terms in this expectation can be evaluated using the same picture we gave for the expectation. Each pair ρ_{ij}, ρ^*_{ij} creates a cupcap matching a pair of indices in one of the first two products with a pair in one of the second two products. However, the isolated edges in C correspond to a fourth-order operator $\mathbb{E}_\sigma(\sigma \otimes \sigma \otimes \sigma^* \otimes \sigma^*)$ which we calculate in the following lemma.

Lemma 7. *If σ is chosen according to the Gaussian measure, then*

$$\mathbb{E}_\sigma \left[\sigma \otimes \sigma \otimes \sigma^* \otimes \sigma^*\right]^{ikmp}_{j\ell nq} = \frac{1}{d^2} \left(\delta^{im}\delta_{jn}\delta^{kp}\delta_{\ell q} + \delta^{ip}\delta_{jq}\delta^{km}\delta_{\ell n} \right) , \tag{24}$$

or diagrammatically,

$$\mathbb{E}_\sigma \left[\sigma \otimes \sigma \otimes \sigma^* \otimes \sigma^*\right] = \frac{1}{d^2} \left(\text{⌒⌒} + \text{⌒⌒} \right) . \tag{25}$$

If σ is chosen according to the Haar measure, then

$$\frac{1 - O(1/d)}{d^2} \left(\text{⌒⌒} + \text{⌒⌒} \right) \preceq \mathbb{E}_\sigma \left[\sigma \otimes \sigma \otimes \sigma^* \otimes \sigma^*\right] \preceq \frac{1 + O(1/d)}{d^2} \left(\text{⌒⌒} + \text{⌒⌒} \right) , \tag{26}$$

where we write $A \preceq B$ if $B - A$ is positive semidefinite.

Proof. We have

$$\mathbb{E}_\sigma \left[\sigma \otimes \sigma \otimes \sigma^* \otimes \sigma^*\right]^{ikmp}_{j\ell nq} = \mathbb{E} \left[\sigma^i_j \sigma^k_\ell (\sigma^m_n)^* (\sigma^p_q)^* \right] .$$

In the Gaussian measure, if $i = m$, $j = n$, $k = p$, and $\ell = q$, but $i \neq k$ or $j \neq \ell$, this gives $\left|\sigma^i_j\right|^2 \left|\sigma^k_\ell\right|^2 = 1/d^2$. If $i = p$, $j = q$, $k = m$, and $\ell = n$, but $i \neq k$ or $j \neq \ell$, we get the same result. Finally, if $i = k = m = p$ and $j = \ell = n = p$, we get $\mathbb{E}\left|\sigma^i_j\right|^4 = 2/d^2$.

In the Haar measure, analogous to Lemma 4 we will calculate the expectation of $\sigma \otimes \sigma \otimes \sigma^* \otimes \sigma^*$ by considering tensor powers of the defining representation σ of $\mathsf{U}(d)$. The tensor square $\sigma \otimes \sigma$ decomposes into symmetric and antisymmetric subspaces, each of which is irreducible:

$$\sigma \otimes \sigma = \tau_+ \oplus \tau_- .$$

The dimension of $\tau_\pm$ is $d_\pm = (d^2 \pm d)/2$. We can write the projection operators onto $\tau_\pm$ in terms of the exchange operator ⨯ which reverses the order of the tensor product, and the identity | | :

$$\Pi_\pm = \frac{1}{2} \left(\text{| |} \pm \text{⨯} \right) .$$

Now writing $\sigma \otimes \sigma \otimes \sigma^* \otimes \sigma^* = (\tau_+ \oplus \tau_-) \otimes (\tau_+^* \oplus \tau_-^*)$, the expectation over σ is the projection operator onto the trivial subspaces of $\tau_+ \otimes \tau_+^*$ and $\tau_- \otimes \tau_-^*$:

$$\mathbb{E}_\sigma \left[\sigma \otimes \sigma \otimes \sigma^* \otimes \sigma^*\right] = \Pi_{\mathbb{1}}^{\tau_+ \otimes \tau_+^*} \oplus \Pi_{\mathbb{1}}^{\tau_- \otimes \tau_-^*} . \tag{27}$$

Analogous to (13), we have the handsome

$$\Pi_{\mathbb{1}}^{\tau_+ \otimes \tau_+^*} = (\Pi_+ \otimes \Pi_+) \cdot \left(\frac{1}{d_+} \right) \cdot (\Pi_+ \otimes \Pi_+) \tag{28}$$

and similarly for $\Pi_{\mathbb{1}}^{\tau_- \otimes \tau_-^*}$. Putting these diagrams together with (27) gives

$$\begin{aligned}
\mathbb{E}_\sigma \left[\sigma \otimes \sigma \otimes \sigma^* \otimes \sigma^*\right] &= (\Pi_+ \otimes \Pi_+) \cdot \left(\frac{1}{d_+} \right) \cdot (\Pi_+ \otimes \Pi_+) + (\Pi_- \otimes \Pi_-) \cdot \left(\frac{1}{d_-} \right) \cdot (\Pi_- \otimes \Pi_-) \\
&= \frac{1}{4d_+} \left(+ + + \right) + \frac{1}{4d_-} \left(- - - + \right) \\
&= \frac{1}{d^2 - 1} \left(+ - \frac{1}{d} \left(+ \right) \right) .
\end{aligned} \tag{29}$$

One can check that (29) is the projection operator onto the two-dimensional subspace spanned by the images of

$$\quad \text{and} \quad ,$$

that is, the vectors $\mathbf{u} = \frac{1}{d} \sum_{i,j} (i,j,i,j)$ and $\mathbf{v} = \frac{1}{d} \sum_{i,j} (i,j,j,i)$. In general, given two real-valued vectors $\mathbf{u}$ and $\mathbf{v}$ of norm 1, let $\Pi_{\mathbf{u}}$ and $\Pi_{\mathbf{v}}$ denote the projection operators onto the subspaces parallel to them, and let $\Pi_{\mathbf{u},\mathbf{v}}$ be the projection operator onto the two-dimensional subspace they span. Then

$$\frac{1}{1 + |\langle \mathbf{u}, \mathbf{v} \rangle|} (\Pi_{\mathbf{u}} + \Pi_{\mathbf{v}}) \preceq \Pi_{\mathbf{u},\mathbf{v}} \preceq \frac{1}{1 - |\langle \mathbf{u}, \mathbf{v} \rangle|} (\Pi_{\mathbf{u}} + \Pi_{\mathbf{v}}) ,$$

where we write $A \preceq B$ if $B - A$ is positive semidefinite. To see this, note that the eigenvectors of $\Pi_{\mathbf{u}} + \Pi_{\mathbf{v}}$ are $\mathbf{u} \pm \mathbf{v}$, with eigenvalues $\lambda_\pm = 1 \pm \langle \mathbf{u}, \mathbf{v} \rangle$, while their eigenvalues with respect to $\Pi_{\mathbf{u},\mathbf{v}}$ are 1. In this case, we have $\langle \mathbf{u}, \mathbf{v} \rangle = 1/d$ and

$$\Pi_{\mathbf{u}} = \frac{1}{d^2} \quad \text{and} \quad \Pi_{\mathbf{v}} = \frac{1}{d^2} .$$

Thus (29) becomes

$$\left(\frac{1}{1 + 1/d} \right) \frac{1}{d^2} \left(+ \right) \preceq \mathbb{E}_\sigma \left[\sigma \otimes \sigma \otimes \sigma^* \otimes \sigma^*\right] \preceq \left(\frac{1}{1 - 1/d} \right) \frac{1}{d^2} \left(+ \right) ,$$

completing the proof. □

The operator corresponds to the coloring $(\kappa, \mu), (\lambda, \nu)$, in which some ρ_{ij} appears in the first and third products, and another ρ'_{ij} appears in the second and fourth. Similarly, the operator corresponds to the coloring $(\kappa, \nu), (\lambda, \mu)$, in which ρ_{ij} appears in the first and fourth products and ρ'_{ij} appears in the second and third. Thus Lemma 7 tells us that, with a multiplicative cost of $1 + O(1/d)$ per isolated edge in the Haar measure, we can replace a given isolated edge in C with

an (unordered) pair of edges. This pair can be colored in two ways: with (κ, μ) and (λ, ν), or with (κ, ν) and (λ, μ). Equivalently, we can "decouple" each quadruple product $\rho \otimes \rho \otimes \rho^* \otimes \rho^*$ into the sum of two combinations of tensor products,

$$\rho \otimes \rho \otimes \rho^* \otimes \rho^* \approx \rho' \otimes \rho'' \otimes \rho'^* \otimes \rho''^* + \rho' \otimes \rho'' \otimes \rho''^* \otimes \rho'^* \,, \tag{30}$$

where ρ' and ρ'' are chosen independently.

Next we explore the set of $(\kappa, \lambda, \mu, \nu)$ corresponding to a given C, or equivalently the set of colorings of C. We will call a coloring *pure* if every edge in C are colored (κ, μ) or (λ, ν). This corresponds to pairing the ρ_{ij}s in the first product in (23) with their conjugates in the third, and those in the second product with their conjugates in the fourth—and choosing the first term in (30) for each ρ_{ij} which appears in all four products. Each cycle in C has two pure colorings, and each isolated edge has one. Thus the number of pure colorings of C is $2^{t(C)}$ where $t(C)$ is the number of cycles in C. A well-known bijection shows that $(\operatorname{perm} A)^2$ can be written as a sum over cycle covers of the bipartite graph defined by A,

$$(\operatorname{perm} A)^2 = \sum_{C \vdash A} 2^{t(C)} \,, \tag{31}$$

or equivalently that $(\operatorname{perm} A)^2$ is the total number of pure colorings. Combining this with (9), we have

$$\mathbb{E}[X]^2 = \sum_{C \vdash A} \sum_{\substack{(\kappa,\lambda,\mu,\nu) \vdash C \\ \text{pure}}} 1 \,. \tag{32}$$

On the other hand, we can associate each coloring with a pure one, say by replacing the color (κ, ν) with (κ, μ) and (λ, μ) with (λ, ν) on each edge. If this converts a tuple of permutations $(\kappa, \lambda, \mu', \nu')$ to a tuple $(\kappa, \lambda, \mu, \nu)$ corresponding to a pure coloring, we will write $(\mu', \nu') \vdash (\kappa, \lambda, \mu, \nu)$. Then, at the risk of some notational overload, we write (23) as a sum over pure colorings:

$$\mathbb{E}[X^2] = \sum_{C \vdash A} \sum_{\substack{(\kappa,\lambda,\mu,\nu) \vdash C \\ \text{pure}}} \sum_{(\mu',\nu') \vdash (\kappa,\lambda,\mu,\nu)} \mathbb{E}_{\{\rho_{ij}\}} \left(\mathbf{tr} \prod_i \rho_{i,\kappa i} \right) \left(\mathbf{tr} \prod_i \rho_{i,\lambda i} \right) \left(\mathbf{tr} \prod_i \rho^*_{i,\mu i} \right) \left(\mathbf{tr} \prod_i \rho^*_{i,\nu i} \right).$$

Now, analogous to [KKL+93], we bound the critical ratio $\mathbb{E}[X^2] / \mathbb{E}[X]^2$ as the maximum ratio between corresponding terms in these two sums, associated with some pure coloring of some cycle cover. The worst possible case is when C consists entirely of isolated edges, since in that case we can switch the colors on each edge independently, giving 2^n colorings for the single pure one.

We can parametrize these 2^n colorings by strings $s \in \{0, 1\}^n$, where $s_i = 0$ if the coloring of the ith edge is pure, and 1 if its colors are switched. This produces diagrams such as those shown in Fig. 2, weaving a total of $8n$ vertices together. As in our calculation of the expectation, the corresponding product of traces is d^{c-2n} where c is the number of loops in this diagram.

Both the pure and "completely impure" colorings 0^n and 1^n—where the ρs in the first product are all paired with those in the third or fourth respectively, and the those in the second product are all paired with those in the fourth or third—have $2n$ loops. In general, the number of loops is $2n$ minus the number of times s switches back and forth between 0 and 1 when s is arranged cyclically. Specifically, there are two loops of length 4 for each i where $s_i = s_{(i+1) \bmod n}$, and a cycle of length 8 for each i where $s_i \neq s_{(i+1) \bmod n}$.

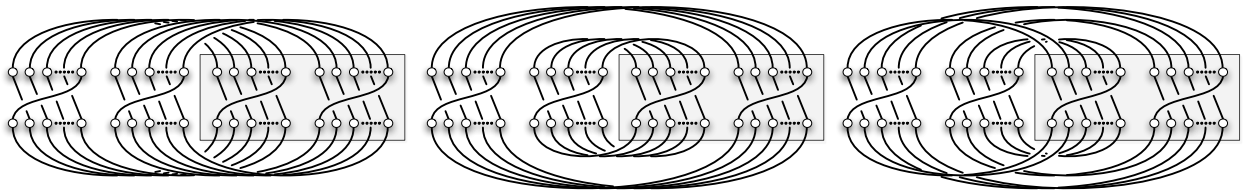


Figure 2: Terms corresponding to a given cycle cover C, where the ρs in the gray box are conjugated and $n = 5$. Left, a pure coloring, which has $2n$ loops. Middle, a maximally impure coloring, which also has $2n$ loops. Right, a mixed coloring corresponding to the string $s = 00111$. A careful inspection shows that it has 8 loops: 6 of length 4, and 2 of length 8.

For each even i with $0 \le i \le n$, there are $2\binom{n}{i}$ strings which switch back and forth i times. Therefore, combined with Lemma 7, we have (for a cycle cover C consisting of n isolated edges)

$$\sum_{(\mu',\nu') \vdash (\kappa,\lambda,\mu,\nu)} \mathbb{E}_{\{\rho_{ij}\}} \left(\mathbf{tr} \prod_i \rho_{i,\kappa i}\right) \left(\mathbf{tr} \prod_i \rho_{i,\lambda i}\right) \left(\mathbf{tr} \prod_i \rho^*_{i,\mu i}\right) \left(\mathbf{tr} \prod_i \rho^*_{i,\nu i}\right)$$
$$= \left(1 + O\left(\frac{1}{d}\right)\right)^n \times 2 \sum_{i=0,2,4,\ldots}^{n} \binom{n}{i} d^{-i} = \left(1 + O\left(\frac{1}{d}\right)\right)^n \times \left(\left(1 + \frac{1}{d}\right)^n + \left(1 - \frac{1}{d}\right)^n\right)$$

In the Gaussian measure, this expression is exact if we remove the prefactor $(1 + O(1/d))^n$; but in any case, we get a bound $(1 + O(1/d))^n$ in either measure. Combining this with (32) completes the first part of the proof of Theorem 1.

4 The second moment in the symmetrized case

Our analysis of the second moment in the symmetrized case proceeds in two steps. We begin, as with the unsymmetrized case, by diagrammatically analyzing the relevant traces. The result is a sum over double cycle covers weighted by an exponential generating function $\sum_\pi d^{c(\pi)}$ over a subset of the symmetric group S_{2n}. We then show that an allied quantity can be analyzed, as in Lemma 5, by harmonic analysis on S_{2n}.

Before stating the main lemmas of this section, we introduce some further notation. As in (31), $t(C)$ denotes the number of cycles in C. As before, we let r denote the rotation $(1, 2, \ldots, n) \in S_n$. The expression $\pi^\sigma = \sigma^{-1}\pi\sigma$ denotes conjugation, and, for two elements $\pi, \sigma \in S_n$, we let (π, σ) denote the element of S_{2n} given by applying π and σ to the first n and last n elements of $\{1, \ldots, 2n\}$, respectively. Finally, we let w_k denote the involution $(1\ n+1)(2\ n+2)\cdots(k\ n+k)$ with the convention that w_0 is the identity. We can then write $\mathbb{E}[X_s^2]$ in terms of the following quantity:

$$a_d^{(2)} = \sum_{k=0}^{n} \binom{n}{k} \frac{1}{d^{2n}} \mathbb{E}_{\alpha,\beta,\gamma,\delta}\, d^{c\left((r^{-1},r^{-1})^{(\alpha,\beta)} w_k (r,r)^{(\gamma,\delta)} w_k\right)}\,.$$

Lemma 8. *If the ρ_{ij} are drawn according to the Gaussian or Haar measure,*

$$\frac{\mathbb{E}[X_s^2]}{\mathbb{E}[X_s]^2} \le \left(1 + O\left(\frac{1}{d}\right)\right)^n \frac{a_d^{(2)}}{a_d^2}\,.$$

We delay the proof of Lemma 8 just long enough for some comforting words regarding the major remaining obstacle: estimating $a_d^{(2)}$. While we do not have a simple, exact expression for $a_d^{(2)}$, we *can* control a larger quantity,

$$\tilde{a}_d^{(2)} = \sum_{k=0}^{n} \binom{n}{k}^2 \frac{1}{d^{2n}} \, \mathbb{E}_{\alpha,\beta,\gamma,\delta} \, d^{c\left((r^{-1},r^{-1})^{(\alpha,\beta)} w_k (r,r)^{(\gamma,\delta)} w_k\right)} \,,$$

in which the kth term of the sum is graced with an extra factor of $\binom{n}{k}$. With this reweighting we can analyze $\tilde{a}_d^{(2)}$ in terms of the Fourier expansions of the class function $d^{c(\cdot)}$, determined by the Kostka numbers, and the convolution square of the conjugacy class $\{(r,r)^\sigma \mid \sigma \in S_{2n}\}$, determined by the Murnaghan-Nakayama rule. This results in the following bound.

Lemma 9. *With notation as above,*

$$\frac{1}{\binom{n}{n/2}} \cdot \tilde{a}_d^{(2)} \;\le\; a_d^{(2)} \;\le\; \tilde{a}_d^{(2)}$$

and

$$\frac{1}{d^{2n}} \binom{2n}{n} \binom{2n+d-1}{2n} \;\le\; \tilde{a}_d^{(2)} \;\le\; \frac{4n^2}{d^{2n}} \binom{2n}{n} \binom{2n+d-1}{2n} .$$

Combining this with Lemmas 8 and 5 completes the proof of (4) and (5) in Theorem 1.

We return now to the proofs of these two lemmas.

Proof of Lemma 8. Squaring (8), the second moment of the symmetrized estimator can be written

$$\mathbb{E}[X_s^2] = \sum_C \sum_{(\kappa,\lambda,\mu,\nu)\vdash C} \mathbb{E}_{\alpha,\beta,\gamma,\delta} \, \mathbb{E}_{\{\rho_{ij}\}} \left(\mathbf{tr} \prod_i \rho_{\alpha i,\kappa\alpha i} \right) \left(\mathbf{tr} \prod_i \rho_{\beta i,\lambda\beta i} \right) \left(\mathbf{tr} \prod_i \rho^*_{\gamma i,\mu\gamma i} \right) \left(\mathbf{tr} \prod_i \rho^*_{\delta i,\nu\delta i} \right) . \tag{33}$$

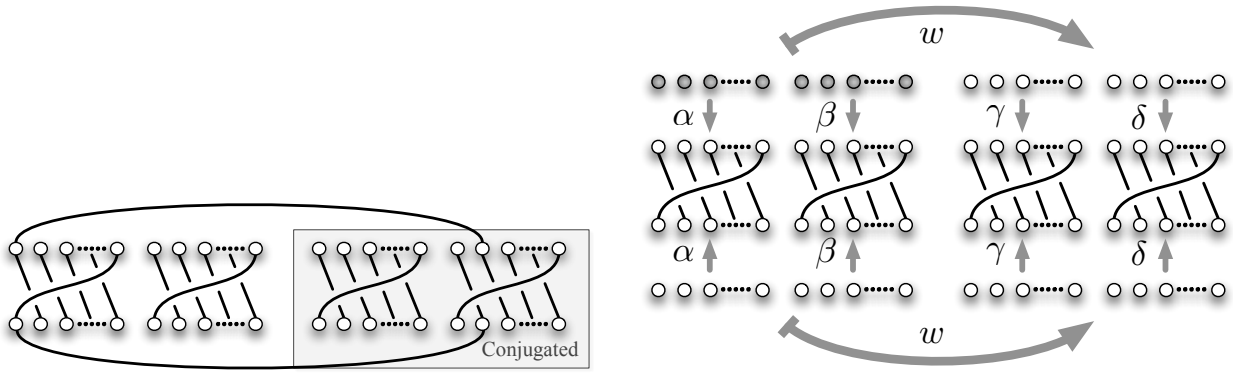
Consider now a term of (33) corresponding to a tuple (κ,λ,μ,ν) of the form

$$\mathbb{E}_{\alpha,\beta,\gamma,\delta} \left(\mathbf{tr} \prod_i \rho_{\alpha i,\kappa\alpha i} \right) \left(\mathbf{tr} \prod_i \rho_{\beta i,\lambda\beta i} \right) \left(\mathbf{tr} \prod_i \rho^*_{\gamma i,\mu\gamma i} \right) \left(\mathbf{tr} \prod_i \rho^*_{\delta i,\nu\delta i} \right) . \tag{34}$$

In light of Lemma 7 (cf. (30)), we may "decouple" any four appearances of the same ρ_{ij}, resulting in a sum of terms in which no ρ appears more than twice. For this reason, we begin our analysis with the extra assumption that each ρ_{ij} appears exactly twice. For notational convenience, let us temporarily refer to the $2n$ distinct ρ_{ij} appearing in (34) simply by $\rho_1, \rho_2, \ldots, \rho_{2n}$, this list in the natural order given by κ and λ (e.g., $\rho_i = \rho_{i,\kappa i}$ and $\rho_{n+i} = \rho_{i,\lambda i}$ for $i \le n$). For a tuple $(\alpha,\beta,\gamma,\delta)$, then, the cupcaps of Eq. (13) introduce edges between conjugate appearances of the same ρ_i as shown in Figure 3(a); any two indices attached by an edge are constrained to be equal.

With this convention, the permutations μ and ν determine a permutation $w \in S_{2n}$ given by the ordering of the conjugate appearances of the ρ_i (when $\alpha = \beta = \gamma = \delta = 1$). The contraction determined by w and a particular $(\alpha,\beta,\gamma,\delta)$ is combinatorial in the sense that it merely constrains families of indices (among the $[\rho_i]_s^t$ and their conjugates) to be equal. Recalling that each cupcap contributes a factor of $1/d$ and each cycle permits d different settings of the indices it contains, the value of this contraction is determined by the cycle structure of the permutation

$$(r^{-1},r^{-1})^{(\alpha^{-1},\beta^{-1})} \; w^{-1} \; (r,r)^{(\gamma,\delta)} \; w \,;$$



(a) Cupcaps and rotations

(b) Symmetrization induces conjugation

Figure 3: Contractions in the second moment computation

see Figure 3(b). In particular, we may write the quantity of (34) as

$$\frac{1}{d^{2n}}\,\mathbb{E}_{\alpha,\beta,\gamma,\delta}\, d^{c\left((r^{-1},r^{-1})^{(\alpha,\beta)}w^{-1}(r,r)^{(\gamma,\delta)}w\right)},$$

where, as before, $c(\pi)$ denotes the number of cycles in the permutation π.

As we are interested in the expectation, over all rearrangements determined by α, β, γ, and δ, the only relevant feature of the permutation w is

$$k = k_{\kappa,\nu} = \Big|w\big(\{1,\ldots,n\}\cap\{n+1,\ldots,2n\}\big)\Big| = \Big|\big\{(i,\kappa(i))\big\}\cap\big\{(i,\nu(i))\big\}\Big|, \tag{35}$$

the number of σ_i carried from the "κ-block" to the "ν-block." Defining $w_k = (1\ n+1)\cdots(k\ n+k)$, we may rewrite the expectation of (34) as

$$\frac{1}{d^{2n}}\,\mathbb{E}_{\alpha,\beta,\gamma,\delta}\, d^{c\left((r^{-1},r^{-1})^{(\alpha,\beta)}w_k(r,r)^{(\gamma,\delta)}w_k\right)}.$$

As in Section 3, for a given double cycle cover C, a coloring $(\kappa,\lambda,\mu,\nu)\vdash C$ is determined by selecting, for each nontrivial cycle c of C, whether c's colors alternate between (κ,μ) and (λ,ν) or (κ,ν) and (λ,μ), and the parity of this coloring. In light of the decoupling equation (30), we may treat each isolated edge as an "unordered pair" of edges that can be colored in two possible ways, with (κ,μ) and (λ,ν) or (κ,ν) and (λ,μ). Recall that in the case of Haar measure, this introduces a factor $1+O(1/d)$ for each isolated edge, giving the factor $(1+O(1/d))^n$.

Observe now that the value of k determined in Eq. (35) is unaffected by the choice of parity in a nontrivial cycle. The other choices described above (determining the colors involved in a nontrivial cycle or isolated edge) have the effect of exchanging a family of ρ_{ij} in the μ-block with a family in the ν-block.

In particular, focusing on the portion of the second moment corresponding to a particular double

cycle cover C, we have

$$\sum_{(\kappa,\lambda,\mu,\nu)\vdash C} \mathbb{E}_{\alpha,\beta,\gamma,\delta} \left(\mathbf{tr}\prod_i \rho_{\alpha i,\kappa\alpha i}\right)\left(\mathbf{tr}\prod_i \rho_{\beta i,\lambda\beta i}\right)\left(\mathbf{tr}\prod_i \rho^*_{\gamma i,\mu\gamma i}\right)\left(\mathbf{tr}\prod_i \rho^*_{\delta i,\nu\delta i}\right)$$

$$= \sum_{(\kappa,\lambda,\mu,\nu)\vdash C} \mathbb{E}_{\alpha,\beta,\gamma,\delta}\, d^{c\left((r^{-1},r^{-1})^{(\alpha,\beta)}w_{\kappa,\nu}(r,r)^{(\gamma,\delta)}w_{\kappa,\nu}\right)} \tag{36}$$

$$\leq \left(1+O\left(\frac{1}{d}\right)\right)^n 2^{t(C)} \sum_{k=0}^{n} \binom{n}{k} \frac{1}{d^{2n}}\, \mathbb{E}_{\alpha,\beta,\gamma,\delta}\, d^{c\left((r^{-1},r^{-1})^{(\alpha,\beta)}w_k(r,r)^{(\gamma,\delta)}w_k\right)} \tag{37}$$

$$= \left(1+O\left(\frac{1}{d}\right)\right)^n 2^{t(C)} a_d^{(2)}\,. \tag{38}$$

Summing over all cycle covers $C \vdash A$ and applying (31) completes the proof. For the Gaussian measure, the same proof applies without the factor $(1+O(1/d))^n$. □

We return to the proof of Lemma 9.

Proof of Lemma 9. The inequality

$$\frac{1}{\binom{n}{n/2}}\tilde{a}_d^{(2)} \;\leq\; a_d^{(2)} \;\leq\; \tilde{a}_d^{(2)}$$

is immediate from the fact that the terms of the sums defining these quantities are positive. We introduce some further notation: for a permutation $\pi \in S_{2n}$, we define

$$\pi^{\uparrow} = \left\{i \mid i \in \{1,\dots,n\}, \pi i \in \{n{+}1,\dots,2n\}\right\} \quad \text{and} \quad \pi^{\downarrow} = \left\{i \mid i \in \{n{+}1,\dots,2n\}, \pi i \in \{1,\dots,n\}\right\}.$$

Then $|\pi^{\uparrow}| = |\pi^{\downarrow}|$ and, if π is selected uniformly in S_{2n}, $\Pr\left[|\pi^{\uparrow}| = k\right] = \binom{n}{k}^2 / \binom{2n}{n}$. Observe also that if α, β, γ, and δ are chosen uniformly from S_n, the element $(\gamma,\delta)w_k(\alpha,\beta)$ is uniform in the set $\{\pi \mid |\pi^{\uparrow}| = k\}$. Recalling that $d^{c(\cdot)}$ is a class function,

$$\frac{1}{\binom{2n}{n}} \cdot \tilde{a}_d^{(2)} = \frac{1}{\binom{2n}{n}} \sum_k \binom{n}{k}^2 \frac{1}{d^{2n}}\, \mathbb{E}_{\alpha,\beta,\gamma,\delta}\, d^{c\left((r^{-1},r^{-1})^{(\alpha,\beta)}w_k(r,r)^{(\gamma,\delta)}w_k\right)} \tag{39}$$

$$= \frac{1}{\binom{2n}{n}} \sum_k \binom{n}{k}^2 \frac{1}{d^{2n}}\, \mathbb{E}_{\alpha,\beta,\gamma,\delta}\, d^{c\left((r^{-1},r^{-1})(\alpha,\beta)^{-1}w_k(\gamma,\delta)^{-1}(r,r)(\gamma,\delta)w_k(\alpha,\beta)\right)}$$

$$= \frac{1}{d^{2n}}\, \mathbb{E}_{\pi}\, d^{c\left((r^{-1},r^{-1})(r,r)^{\pi}\right)} = \frac{1}{d^{2n}}\, \mathbb{E}_{\pi}\, \mathbb{E}_{\sigma}\, d^{c((r,r)^{\sigma}(r,r)^{\pi})}\,,$$

where π and σ are chosen uniformly at random from S_{2n}. Here we use the fact that any element of S_{2n}—in this case, (r,r)—is in the same conjugacy class as its inverse.

Defining $P_{n,n}$ to be the uniform distribution on the conjugacy class

$$[(r,r)] = \{(r,r)^{\pi} \mid \pi \in S_{2n}\} \subset S_{2n}\,,$$

we may express the quantity above as an inner product

$$\frac{1}{\binom{2n}{n}}\tilde{a}_d^{(2)} = \frac{1}{d^{2n}}(2n)!\, \left\langle d^{c(\cdot)}, P_{n,n} * P_{n,n} \right\rangle. \tag{40}$$

As in the proof of Lemma 5, we compute this inner product by determining the Fourier expansions of the class functions $d^{c(\cdot)}$ and $P_{n,n}$. By the Murnaghan-Nakayama rule, $\chi_\lambda(n,n) = 0$ unless the tableau λ can be expressed as the union of two n-ribbon tiles. Any such tableau has *rank* (the number of cells on the diagonal) no more than two and can be conveniently expressed in terms of its *characteristics*: defining a_i and b_i to be the number of cells below and to the right of the ith box of the diagonal, respectively, we use the notation $\tau = (b_1, b_2, \ldots, b_r \mid a_1, a_2, \ldots, b_r)$ to describe the tableau (see Figure 4). If $\chi_\tau(n,n)$ is nonzero, so that τ can be written as the union of two

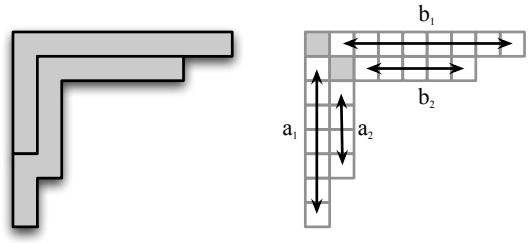


Figure 4: A Young tableau decomposed into two n-ribbon tiles.

n-ribbons, we find (again appealing to the Murnaghan-Nakayama rule) that either

- $\tau = (b_1, b_2 \mid a_1, a_2)$ has rank two, $a_1 + b_2 + 1 = a_2 + b_1 + 1 = n$, and $\chi_\tau(n,n) = \pm 2$, or
- $\tau = (b_1 \mid a_1)$ has rank one and $\chi_\tau(n,n) = \pm 1$.

We let T_n denote the family of representations of S_{2n} described above; note that $|T_n| \leq n^2$. Observe that for each $\tau \in T_n$, $\langle P_{n,n}, \chi_\tau \rangle = \frac{1}{(2n)!}\chi_\tau(n,n)$ (where $\chi_\tau(n,n) \in \{\pm 1, \pm 2\}$). Recalling that

$$\chi * \chi = \frac{|G|}{\chi(1)}\chi$$

for any irreducible character χ of a group G, we may express

$$\langle P_{n,n} * P_{n,n}, \chi_\tau \rangle = \frac{1}{(2n)!}\frac{\chi_\tau(n,n)^2}{\dim \tau}\,.$$

As discussed in the proof of Lemma 5,

$$\left\langle d^{c(\cdot)}, \chi_\tau \right\rangle = \langle \chi_\Sigma, \chi_\tau \rangle = \sum_{\substack{(\rho_1,\ldots,\rho_d) \\ \sum \rho_i = 2n}} K_\rho^\tau\,,$$

where χ_Σ is the permutation representation given by the action of S_{2n} on the set $\{(a_1, \ldots, a_{2n} \mid a_i \in \{1, \ldots, d\}\}$ and K_ρ^τ is the Kostka number, equal to the number of semistandard tableaux of shape τ with ρ_i appearances of the number i. Then

$$\langle P_{n,n} * P_{n,n}, d^{c(\cdot)} \rangle = \sum_\tau \langle P_{n,n} * P_{n,n}, \chi_\tau \rangle \left\langle d^{c(\cdot)}, \chi_\tau \right\rangle = \frac{1}{(2n)!}\sum_{\tau \in T_n} \chi_\tau(n,n)^2 \sum_{\substack{(\rho_1,\ldots,\rho_d) \\ \sum \rho_i = 2n}} \frac{K_\rho^\tau}{\dim \tau}\,. \quad (41)$$

Note that for each $\tau \in T_n$, $\chi_\tau(n,n)^2 \le 4$ and $K_\rho^\tau \le \dim \tau$, as $\dim \tau$ is the number of semistandard tableaux of shape τ with distinct entries in any totally ordered set. Thus,

$$\langle P_{n,n} * P_{n,n}, d^{c(\cdot)} \rangle \le \frac{4}{(2n)!} \sum_{\tau \in T_n} \binom{2n+d-1}{2n} \le \frac{4}{(2n)!} |T_n| \binom{2n+d-1}{2n} \le \frac{4n^2}{(2n)!} \binom{2n+d-1}{2n}.$$

On the other hand, each term in the sum of (41) is positive; thus

$$\langle P_{n,n} * P_{n,n}, d^{c(\cdot)} \rangle \ge \langle P_{n,n} * P_{n,n}, \chi_1 \rangle \left\langle d^{c(\cdot)}, \chi_1 \right\rangle = \frac{1}{2n!} \binom{2n+d-1}{2n}. \tag{42}$$

We conclude that

$$\frac{1}{2n!} \binom{2n+d-1}{2n} \le \langle P_{n,n} * P_{n,n}, d^{c(\cdot)} \rangle \le \frac{4n^2}{2n!} \binom{2n+d-1}{2n}$$

which, in conjunction with (40), completes the proof of Lemma 9. □

Now we apply these Lemmas to prove an upper bound on the critical ratio $\mathbb{E}[X_s^2]/\mathbb{E}[X_s]^2$. If d is constant, which is the only case for which we have an efficient algorithm to compute X_s [Bar00], our bound is not very inspiring. If $n \ge d$, combining Lemmas 5, 8, and 9 gives

$$\frac{\mathbb{E}[X_s^2]}{\mathbb{E}[X_s]^2} \le 4n^2 \binom{2n}{n} \binom{2n+d-1}{2n} \Big/ \binom{n+d}{n+1}^2 = O(n^3/d) \binom{2n+2d}{n+d} \Big/ \binom{2n+2d}{d} = 2^{2n}\, n^{-d+O(1)},$$

assuming that $d = O(1)$. This proves (4) in Theorem 1, and suggests that d needs to grow with n to give a good estimator.

On the other hand, when d grows fast enough with n, we find that the critical ratio behaves quite well. Combining Lemma 9 with the lower bound (15) gives

$$\frac{\mathbb{E}[X_s^2]}{\mathbb{E}[X_s]^2} \le \frac{4n!^2}{d^{2n}} \binom{2n}{n} \binom{2n+d-1}{2n} = \frac{4}{d^{2n}} \frac{(2n+d-1)!}{(d-1)!} = 4\left(1+\frac{1}{d}\right) \cdots \left(1+\frac{2n-1}{d}\right) \le 4\mathrm{e}^{4n^2/d},$$

completing the proof of (5) in Theorem 1.

In the critical case where $d = O(1)$ the upper bound of $2^{2n} n^{-d+O(1)}$ we establish above is tight up to the factor introduced by our "approximation" of $a_d^{(2)}$ by $\tilde{a}_d^{(d)}$—that is, a factor of $\binom{n}{n/2}$. In particular, even for the identity matrix, we can establish a $2^n n^{-d+O(1)}$ lower bound on the critical ratio:

Theorem (Restatement of Theorem 2)**.** *Let A be the $n \times n$ identity matrix and d a constant. Then*

$$\frac{\mathbb{E}[X_s^2]}{\mathbb{E}[X_s]^2} = \Omega\left(\frac{2^n}{n^d}\right) \quad \textit{and} \quad \frac{\mathbb{E}[X_s^2]}{\mathbb{E}[X_s]^2} = \left(1 - O\left(\frac{1}{d}\right)\right)^n \Omega\left(\frac{2^n}{n^d}\right),$$

when the ρ_{ij} are distributed according to the Gaussian or Haar measure respectively.

Proof of Theorem 2. Let d be a constant and A the $n \times n$ identity matrix. Then $\operatorname{perm} A = \operatorname{perm}^2 A = 1$ and, from Lemma 5,

$$\mathbb{E}[X_s] = a_d = \frac{1}{d^n} \binom{n+d}{n+1}.$$

As for the second moment, the only nontrivial term in the sum (33) corresponds to the case where the permutations κ, λ, μ, and ν are the identity. In this case each ρ_{ij} appears four times and there are precisely $\binom{n}{k}$ terms of (36) for which $w_{\kappa,\lambda} = w_k$; in particular, in this case the inequality of (37) is an equality. Recalling Lemma 7, we conclude that

$$\mathbb{E}[X_s^2] = a_d^{(2)} \quad \text{and} \quad \mathbb{E}[X_s^2] \geq (1 - O(1/d))^n \, a_d^{(2)}$$

when the ρ_{ij} have Gaussian measure and Haar measure, respectively. For constant d we have

$$\frac{\tilde{a}_d^{(2)}}{a_d^2} \geq \frac{a_d^{(2)}}{\binom{n}{n/2} a_d^2} = \frac{\binom{2n}{n}\binom{2n+d-1}{2n}}{\binom{n}{n/2}\binom{n+d}{n+1}^2}$$

and, considering that $\binom{\ell}{\ell/2} = \frac{2^\ell}{\Theta(\sqrt{\ell})}$ and $\binom{2n+d-1}{2n} \geq \binom{n+d}{n+1}$,

$$\frac{\tilde{a}_d^{(2)}}{a_d^2} = \frac{2^{2n}}{2^n O(\sqrt{n}) \binom{n+d}{n+1}} = \Omega\left(\frac{2^n}{n^d}\right) .$$

The statement of the theorem follows. □

5 Estimators based on the Frobenius norm

In this section, we prove Theorem 3 by relating the moments of Frobenius estimators, $X_{\text{Frob}} = \|\det M\|^2$ and $X_{\text{Frob},s} = \|\text{sdet}\, M\|^2$, to those of the trace-squared estimators we studied above.

As Fig. 5 shows, the diagrams corresponding to the expectations and second moments of these estimators differ from those of their counterparts by a small number of local moves. Let Q be the product of some sequence of ρ_{ij}. Then all we have to do is change our previous contraction,

$$|\mathbf{tr}\, Q|^2 = Q_i^i Q_j^j$$

where the "output" of each Q is connected to its "input," to

$$\|Q\|^2 = \mathbf{tr}\, QQ^\dagger = Q_j^i (Q^\dagger)_i^j = Q_j^i (Q^*)_j^i .$$

In this contraction, we connect the output of each Q to the output of the corresponding Q^*, and similarly wire their inputs together. The cupcaps, resulting from taking the expectation of $\rho \otimes \rho^*$ for each ρ_{ij} appearing in these products, remain the same as before.

Now recall that the expectation and second moment of these estimators is proportional to d^c, where c is the number of loops in these diagrams. Each of these rewiring moves changes the number of loops by at most one, by cutting one loop into two or merging two loops into one. Thus we have

$$\frac{1}{d}\,\mathbb{E}[X] \leq \mathbb{E}[X_{\text{Frob}}] \leq d\;\mathbb{E}[X] \quad \text{and} \quad \frac{1}{d^2}\,\mathbb{E}[X^2] \leq \mathbb{E}[X_{\text{Frob}}^2] \leq d^2\;\mathbb{E}[X^2]\,,$$

and similarly in the symmetrized case. Assuming the worst regarding these bounds yields (6), and completes the proof of Theorem 1.

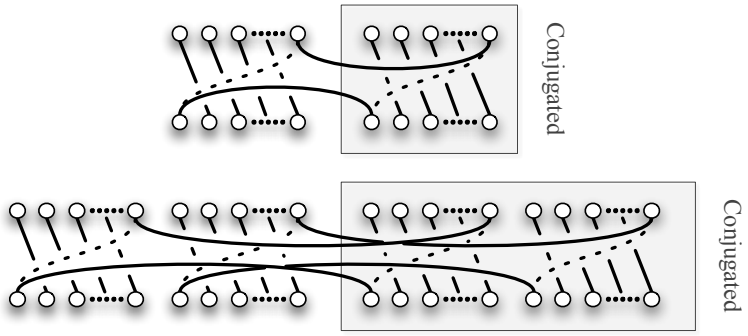


Figure 5: Rewiring the diagram to change $|\mathbf{tr}\, M|^2$ to $\mathbf{tr}\, MM^\dagger = \|M\|^2$. The cupcaps remain unchanged, but instead of wiring the "input" of each product to its "output," we wire a pair of products together "input" to "input" and "output" to "output."

6 Conclusions

As we stated in the Introduction, our results present us with the following irony. For the estimators based on the unsymmetrized determinant, which we do not know how to compute efficiently, the critical ratio $\mathbb{E}[X^2]/\,\mathbb{E}[X]^2$ becomes more mildly exponential as d increases. Specifically, for any $\epsilon > 0$ we can make the critical ratio $O((1+\epsilon)^n)$ by taking $d = 1/\epsilon$.

On the other hand, for the estimators based on the symmetrized determinant, the critical ratio is $\Omega(2^n)$ in the case $d = O(1)$ where we have an efficient algorithm. In order to reduce this exponential to $O(c^n)$ for some $c < 2$, we need d to be a growing function of n. This is contrary to the intuition expressed in [Bar00], and to our own initial intuition when we began work on this problem.

Of course, the symmetrized estimators may still be tightly concentrated, as conjectured in [Bar00]. However, since their variance is large, any proof of concentration would have to bound, implicitly or explicitly, their higher moments.

At this point, finding an algebraic polynomial-time approximation scheme for the permanent seems to require progress on at least one of several fronts. One approach would be to seek a polynomial-time algorithm for sdet M in the case where M's entries belong to $\mathcal{A}_d$ where $d = \mathrm{poly}(n)$, but it seems difficult to scale up the algorithm of [Bar00] beyond $d = O(1)$. Another approach, as suggested in [CRS03], would be to seek an algorithm for $\det M$ where M's entries belong to some group with representations of arbitrarily high dimension. However, it seems difficult to construct a succinct description for the group algebra elements which appear in the determinant, since their support in the group basis is exponentially large.

Acknowledgments

This research was supported by NSF grants CCF-0524613, CCF-0835735, and CCF-0829917.

References

[Bar99] Alexander I. Barvinok. Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor. *Random Structures and Algorithms*, 14(1):29–61, 1999.

[Bar00] Alexander I. Barvinok. New permanent estimators via non-commutative determinants, 2000.

[Con00] John B. Conway. *A Course in Operator Theory*, volume 21 of *Graduate Studies in Mathematics*. American Mathematical Society, 2000.

[CRS03] Steve Chien, Lars Eilstrup Rasmussen, and Alistair Sinclair. Clifford algebras and approximating the permanent. *J. Comput. Syst. Sci.*, 67(2):263–290, 2003.

[GG81] C. D. Godsil and Ivan Gutman. On the matching polynomial of a graph. In *Algebraic Methods in Graph Theory*, pages 241–249. North-Holland, 1981.

[JK81] Gordon James and Adalbert Kerber. *The representation theory of the symmetric group*, volume 16 of *Encyclopedia of mathematics and its applications*. Addison–Wesley, 1981.

[KKL+93] Narendra Karmarkar, Richard M. Karp, Richard J. Lipton, László Lovász, and Michael Luby. A Monte-Carlo algorithm for estimating the permanent. *SIAM J. Comput.*, 22(2):284–293, 1993.

[Nis91] Noam Nisan. Lower bounds for non-commutative computation. In *Proc. 23rd Annual ACM Symposium on Theory of computing*, pages 410–418, New York, NY, USA, 1991. ACM.

[Tod91] Seinosuke Toda. PP is as hard as the polynomial-time hierarchy. *SIAM J. Comput.*, 20(5):865–877, 1991.

[Val79] Leslie G. Valiant. The complexity of computing the permanent. *Theor. Comp. Sci.*, 8:189–201, 1979.

A Representation theory and the symmetric group

We briefly discuss the elements of the representation theory of groups, and of the symmetric groups in particular. Our treatment is primarily for the purposes of setting down notation; we refer the reader to [JK81] for a complete account.

Let G be a finite group. A *representation* ρ of G is a homomorphism $\rho : G \to \mathsf{U}(V)$, where V is a finite-dimensional Hilbert space and $\mathsf{U}(V)$ is the group of unitary operators on V. The *dimension* of ρ, denoted d_ρ, is the dimension of the vector space V. By choosing a basis for V, then, we can identify each $\rho(g)$ with a unitary $d_\rho \times d_\rho$ matrix; these matrices then satisfy $\rho(gh) = \rho(g) \cdot \rho(h)$ for every $g, h \in G$.

Fixing a representation $\rho : G \to \mathsf{U}(V)$, we say that a subspace $W \subset V$ is *invariant* if $\rho(g)W \subset W$ for all $g \in G$. We say ρ is *irreducible* if it has no invariant subspaces other than the trivial space $\{\mathbf{0}\}$ and V. If two representations ρ and σ are the same up to a unitary change of basis, we say that they are *equivalent*. It is a fact that any finite group G has a finite number of distinct irreducible

representations up to equivalence and, for a group G, we let $\hat{G}$ denote a set of representations containing exactly one from each equivalence class. The irreducible representations of G give rise to the Fourier transform. Specifically, for a function $f : G \to \mathbb{C}$ and an element $\rho \in \hat{G}$, define the *Fourier transform of f at ρ* to be

$$\hat{f}(\rho) = \sqrt{\frac{d_\rho}{|G|}} \sum_{g \in G} f(g)\rho(g) \ .$$

The leading coefficients are chosen to make the transform unitary, so that it preserves inner products:

$$\langle f_1, f_2 \rangle = \sum_g f_1^*(g) f_2(g) = \sum_{\rho \in \hat{G}} \mathbf{tr}\Big(\hat{f}_1(\rho)^\dagger \cdot \hat{f}_2(\rho)\Big) \ .$$

In the case when ρ is *not* irreducible, it can be decomposed into a direct sum of irreducible representations, each one of which operates on an invariant subspace. We write $\rho = \sigma_1 \oplus \cdots \oplus \sigma_k$ and, for the σ_i appearing at least once in this decomposition, $\sigma_i \prec \rho$. In general, a given σ can appear multiple times, in the sense that ρ can have an invariant subspace isomorphic to the direct sum of a_σ^ρ copies of σ. In this case a_σ^ρ is called the *multiplicity* of σ in ρ, and we write $\rho = \bigoplus_{\sigma \prec \rho} a_\sigma^\rho \sigma$.

For a representation ρ we define its *character* as the trace $\chi_\rho(g) = \mathbf{tr}\, \rho(g)$. Given an element m, we denote its conjugacy class $[m] = \{g^{-1}mg \mid g \in G\}$. Since the trace is invariant under conjugation, characters are constant on the conjugacy classes, and we write $\chi_\rho([m]) = \chi_\rho(m)$ where m is any element of $[m]$. Characters are a powerful tool for reasoning about the decomposition of reducible representations. In particular, for $\rho, \sigma \in \hat{G}$, we have the orthogonality conditions

$$\langle \chi_\rho, \chi_\sigma \rangle_G = \frac{1}{|G|} \sum_{g \in G} \chi_\rho(g)\chi_\sigma(g)^* = \begin{cases} 1 & \rho = \sigma \ , \\ 0 & \rho \neq \sigma \ . \end{cases}$$

If ρ is reducible, we have $\chi_\rho = \sum_{\sigma \prec \rho} a_\sigma^\rho \chi_{\sigma_i}$, and so the multiplicity a_σ^ρ is given by

$$a_\sigma^\rho = \langle \chi_\rho, \chi_\sigma \rangle_G \ .$$

If ρ is irreducible, *Schur's lemma* asserts that the only matrices which commute with $\rho(g)$ for all g are the scalars, $\{c\mathbb{1} \mid c \in \mathbb{C}\}$. Therefore, for any A we have

$$\frac{1}{|G|} \sum_{g \in G} \rho(g)^\dagger A \rho(g) = \frac{\mathbf{tr}\, A}{d_\rho} \mathbb{1}_{d_\rho} \tag{43}$$

since conjugating this sum by $\rho(g)$ simply permutes its terms.

We specialize now to the case of the symmetric group S_n of permutations of the set $\{1, \ldots, n\}$. The representations of S_n are in one-to-one correspondence with *Young diagrams* or, equivalently, integer partitions $\lambda = (\lambda_1, \lambda_2, \cdots)$ where $\lambda_1 \geq \lambda_2 \geq \cdots$ and $\sum_i \lambda_i = n$. The character of this representation is denoted χ_λ. The *Murnaghan-Nakayama rule* gives a recursive formula for the character χ_λ. In preparation for stating the rule, we define a *ribbon tile* of length k to be a polyomino of k cells, arranged in a path where each step is up or to the right.

Lemma 10 (Murnaghan-Nakayama rule)**.** *Given a Young diagram λ and a permutation π with cycle structure $k_1 \geq k_2 \geq \cdots$, a* consistent tiling *of λ consists of removing a ribbon tile of length k_1*

from the boundary of λ, then one of length k_2, and so on, with the requirement that the remaining part of λ is a Young diagram at each step. Let h_i denote the height of the ribbon tile corresponding to the ith cycle: then

$$\chi_\lambda(\pi) = \sum_T \prod_i (-1)^{h_i+1} \tag{44}$$

where the sum is over all consistent tilings T.

B Proof of Lemma 4

Proof. For the Gaussian measure, this is simply the fact that $(\sigma \otimes \sigma^*)^{ik}_{j\ell} = \sigma^i_j (\sigma^k_\ell)^*$. If $i \neq k$ or $j \neq \ell$, then this is the product of two independent random variables both of whom have expectation zero. If $i = k$ and $j = \ell$, then this is $\left|\sigma^i_j\right|^2$, whose expectation is $1/d$.

For the Haar measure, (12) follows from a little representation theory. (For a brief introduction to representation theory, see Appendix A.) Abusing notation, suppose that σ is the defining representation of the group $\mathsf{U}(d)$ of unitary matrices, i.e., the d-dimensional representation in which unitary matrices act on column vectors in the natural way. Then $\sigma \otimes \sigma^*$ is isomorphic to the conjugation action of $\mathsf{U}(d)$ on $\mathsf{GL}(d)$, the vector space of $d \times d$ matrices. We can decompose this into the direct sum of two invariant subspaces $\sigma \otimes \sigma^* \cong \mathbb{1} \oplus \Gamma$, where $\mathbb{1}$ is the trivial representation, consisting of the scalar matrices, and Γ is the (d^2-1)-dimensional representation consisting of $d \times d$ matrices with zero trace. Both these subspaces are clearly invariant under conjugation, and are, in fact, irreducible. Taking the expectation over $\sigma \in \mathsf{U}(d)$ gives the projection operator $\Pi_{\mathbb{1}}^{\sigma \otimes \sigma^*}$ onto the trivial subspace—that is, the linear operator on the space of matrices which takes a matrix $A = A_{ik}$ and returns a scalar whose trace is $\mathbf{tr}\, A$. We claim that this operator is exactly (12), since

$$\left(\frac{1}{d}\,\delta^{ik}\delta_{j\ell}\right) A_{ik} = \frac{1}{d}\, A^i_i \delta_{j\ell} = \left(\frac{1}{d}\ \mathbf{tr}\, A\right) \mathbb{1}\,.$$

Here we again use the Einstein summation convention, so that $A^i_i = \mathbf{tr}\, A$, and the identity matrix is $\mathbb{1} = \delta_{j\ell}$. □